

# Quantitative single-cell splicing analysis reveals an ‘economy of scale’ filter for gene expression

Fangyuan Ding<sup>1</sup>, Michael B. Elowitz<sup>1\*</sup>

<sup>1</sup>Division of Biology and Biological Engineering, Caltech, 1200 E. California Blvd. Pasadena, CA 91125, USA

\*melowitz@caltech.edu

## Abstract

In eukaryotic cells, splicing affects the fate of each pre-mRNA transcript, helping to determine whether it is ultimately processed into an mRNA, or degraded. The efficiency of splicing plays a key role in gene expression. However, because it depends on the levels of multiple isoforms at the same transcriptional active site (TAS) in the same cell, splicing efficiency has been challenging to measure. Here, we introduce a quantitative single-molecule FISH-based method that enables determination of the absolute abundances of distinct RNA isoforms at individual TASs. Using this method, we discovered that splicing efficiency behaves in an unexpected ‘economy of scale’ manner, increasing, rather than decreasing, with gene expression levels, opposite to a standard enzymatic process. This behavior could result from an observed correlation between splicing efficiency and spatial proximity to nuclear speckles. Economy of scale splicing represents a non-linear filter that amplifies the expression of genes when they are more strongly transcribed. This method will help to reveal the roles of splicing in the quantitative control of gene expression.

## Introduction

In eukaryotes, most genes undergo co-transcriptional splicing<sup>1–3</sup>. During splicing, nascent pre-mRNAs are processed to remove introns and to include or exclude exons<sup>4</sup>, generating multiple isoforms, with distinct fates, including functional mRNAs and degradation products<sup>5–7</sup>. The ubiquitousness of splicing, and its tight coupling with transcription<sup>1–3</sup> suggest that splicing could play additional roles beyond diversification of protein products per se. In particular, it could act as a signal processing filter, modulating the amount of mature mRNA depending on the rate of transcription, or other parameters. In general, only a fraction of transcribed RNA is productively spliced to enable subsequent translation or other functions. Unspliced RNA is predominantly retained in the nucleus and degraded, through a quality control mechanism<sup>8,9</sup>. A critical feature of splicing is its efficiency, defined as the ratio of spliced RNA to total transcribed RNA (Figure 1a, SI text). In general, it has remained unclear how splicing efficiency depends on the gene expression level, and therefore what type of filter, if any, splicing provides.

In general, splicing efficiency could behave in three qualitatively distinct ways. The simplest possibility is that splicing efficiency is determined by sequence features and concentrations of splicing machinery, and therefore constant for a given gene in a given cell state. However, as an enzymatic process, splicing could in principle decline in efficiency at high substrate (pre-mRNA) concentrations. This ‘diminishing returns’ behavior would tend to disproportionately suppress processing at higher expression levels relative to lower levels. The final formal possibility, which is not expected under conventional models of splicing, is that splicing efficiency could increase with transcription level, in an ‘economy of scale’ fashion, and thereby disproportionately enhance the expression of more highly expressed genes. Accurate measurements of splicing efficiency across transcription levels are needed to discriminate among these or more complex behaviors.

Several approaches have been used to measure splicing efficiency<sup>10–14</sup>. Genome-wide shot-gun high-throughput sequencing provides a way to compute splicing efficiency by comparing the number of reads at unspliced regions to the total number of reads in constitutive exon regions of the same gene. However, much splicing occurs co-transcriptionally<sup>15</sup> and different RNA species have different lifetimes once released from the TAS. Because they do not discriminate RNAs at the TAS from RNAs at other sites, these methods can distort quantification of splice ratios. To circumvent this problem, more recent studies have performed nascent-RNAseq<sup>16,17</sup> to measure co-transcriptional splicing efficiency at the TAS. These approaches have been powerful and informative. However, they necessarily average over individual cells.

Averaging over a heterogeneous cell population can, by itself, distort splicing efficiency. The key issue is that splicing efficiency is an inherently ratiometric quantity. To determine the mean splicing efficiency across a cell population, one would ideally calculate the ratio of spliced to total transcripts in each cell individually and then average this quantity over the cell population. However, because the mean of a ratio is not, in general, equal to the ratio of a mean, splicing efficiency measured from single cells does not match the population average (Figure 1b). More specifically, population-average measurements systematically underweight contributions from lower expressing cells relative to higher expression cells. For a bursty process like gene expression<sup>18–21</sup>, this effect can be strong.

Several previous studies have sought to analyze splicing efficiency in single cells<sup>19,22</sup>. Pioneering studies engineered binding sites for the MS2 and PP7 RNA-binding proteins to fluorescently label individual transcripts in live cells<sup>23,24</sup>. This approach enabled simultaneous analysis of splicing and transcriptional kinetics in individual cells. However, it cannot be used on endogenous (unmodified) transcripts, and insertion of binding sites could potentially perturb the splicing dynamics.

Here, we report a method for quantitative single-cell measurement of splicing efficiency based on single-molecule fluorescence in-situ hybridization (smFISH). The method measures splicing efficiency at transcriptional active sites in individual cells. In contrast to smFISH methods based

on counting the number of distinct molecules, appearing as fluorescent dots in images<sup>25–27</sup>, here we quantify dot intensity at the TAS. For accurate quantitation, we developed methods for unbiased intensity comparisons between channels and adapted a method from astrophysics for estimating stellar luminosities in crowded star fields<sup>28</sup>.

Contrary to the classic enzyme-substrate Michaelis-Menten model, splicing efficiency increased, rather than decreased, with increasing levels of gene expression, in an ‘economy of scale’ fashion. Increased transcription also correlated with spatial proximity to speckles, suggesting a mechanism for economy of scale based on spatial clustering. A mathematical model based on this observation shows how economy of scale splicing could emerge if enzyme availability increases with substrate (pre-mRNA) concentration. Together, these results enable quantitative analysis of splicing in single cells and reveal a new role for splicing as a gene expression filter.

## Results

### Intron and exon probe smFISH sets identify distinct RNA species

We set out to measure splicing efficiency by quantifying the relative levels of different isoforms at the same TAS, across a range of expression levels. As a model system, we used the RG6 mini-gene, whose splicing behavior was previously characterized using fluorescent proteins<sup>29</sup>. To enable regulation of transcription, we site-specifically integrated the mini-gene under the control of a dox-inducible promoter into HEK293 cells. To measure the splicing efficiency at the TAS, we designed three smFISH probe sets. The intron probe set targeted the spliceable constitutive intron, and thus measured the number of unspliced transcripts. The other two probe sets, denoted Exon1 and Exon2, targeted constitutive exons, measuring the number of total transcripts (Figure 2a). The use of two redundant exon probe sets facilitates subsequent analysis (see below). We cultured cells under standard conditions and then fixed and imaged the cells using all three smFISH probe sets (SI Methods and Materials).

We observed several types of smFISH dots that could be classified by their intensity and probe binding patterns. The first type, representing the TAS, consisted of one or two bright dots per cell that appeared in all three channels in the nucleus (Figure 2a and SI Text). The second type consisted of many scattered dots of lower intensity that appeared in both the nucleus and cytoplasm corresponding to single transcripts (Figure 2a). Using co-localization of different probe sets enabled further classification of dots. All RNAs were labeled by both Exon1 and Exon2 probe sets, both in the cytoplasm and the nucleus. In contrast, unspliced isoforms were labeled in all three channels, but appeared only in the nucleus (Figure 2a). Although we targeted a constitutively spliced intron, we also observed released unspliced molecules, consistent with imperfect splicing efficiencies (Figure 2a, ‘unspliced isoform’). These transcripts could result from a failure to complete splicing prior to transcriptional termination, or from

competition among transcripts for limited levels of splicing machinery<sup>30</sup>. Together, these probe sets allowed identification of multiple distinct molecular species both at and away from the TAS.

## Determining single transcript intensity units

Because individual transcript molecules cannot be spatially resolved at the TAS, we developed an intensity-based transcript-counting procedure to quantify the number of transcripts of each species from fluorescence intensities of each probe set (Figure 2b, Figure S1a and SI Text). Briefly, we first used the Poisson-distributed individual transcript intensities to obtain the single molecule fluorescence intensity calibration units in each color channel, denoted  $q_{E1}$ ,  $q_{E2}$ , and  $q_I$  for the two exon and one intron channels, respectively. Then, we used this calibration to quantify the number of copies of exonic ( $N_{E1}$ ,  $N_{E2}$ ) and intronic ( $N_I$ ) targets at the TAS, in molecular units.

In this procedure, the values of  $q_n$  ( $n = E1, E2, \text{ or } I$ ) are estimated from the distribution of intensities of single-molecule non-TAS smFISH dots in each channel. However, dot identification can be inconsistent between channels due to differences in signal-to-background levels, differences in binding efficiencies between probe sets, and differences in the fluorescence properties of each fluorophore, among other issues. These factors produce systematic differences in sensitivity between channels that distort  $q_n$  quantification.

To address this issue, we developed an unbiased dot identification procedure. We first identified candidate dots using a low (permissive) threshold that captures all foreground smFISH dots as well as some background signal (false positive dots) in each channel, and quantified the integrated intensity of each dot. To correct for fluorescence ‘contamination’ from neighboring dots, we adapted an algorithm from stellar photometry of crowded star fields<sup>28</sup>, which works by iteratively removing fluorescence from neighboring objects (Figure 2c and Figure S1c, SI Text).

We also performed this analysis on negative-control images lacking true smFISH dots to obtain the distribution of background dot intensities. Subtracting the background histogram from the total (foreground + background) histogram generated a corrected foreground dot intensity distribution (Figure 3a, and SI Text). TAS intensity measurements do not require accurate counting of scattered smFISH dots, as in conventional FISH<sup>25</sup>. This approach thus effectively sacrifices precision in individual dot identification to obtain a less biased distribution of smFISH dot intensity in each fluorescent channel. Finally, to obtain the single transcript intensity unit  $q_n$ , we fit the resulting distributions with a continuous analog of the Poisson distribution (Figure 3b, and SI Text).

As an additional consistency check, we used an independent method to identify foreground dots by their co-localization across multiple fluorescence channels. This approach produced similar values for  $q_n$  (Figure 3c, Figure S1d, SI Text). It also revealed that foreground and background dots exhibit overlapping distributions of key properties such as intensity, peak height, and peak width (Figure 3d, and Figure S2, S3, S4, SI Text), further supporting the need for histogram subtraction, above. Taken together, these results provide a simple and general method for accurately estimating the fluorescence intensity at the TAS using precisely calibrated intensity units for single transcripts.

### **Fluorescence intensity increases linearly with number of probes**

Because intensity quantification is critical for estimating the number of molecules at the TAS, we next asked how TAS fluorescence intensity depends on the abundance of bound probes. We designed 42 smFISH probes targeting the housekeeping gene *SDHA*, and mixed them into groups of 18, 24, 30, 36, and 42 probes. We then measured the fluorescence intensity of stained HEK293 cells with each group. Fluorescence dot intensities increased linearly with the number of included probes (Figure 3e). By contrast, a second set of 27 probes targeting *HES1*, which were labeled in a different fluorescent channel and included in all experiments as a fixed control, were constant across each condition, as expected. These results indicate that fluorescence intensity provides a linear readout of probe density at the TAS. This linearity enables one to quantify the number of transcripts ( $N_{E1}$ ,  $N_{E2}$ , and  $N_I$ ) at the TAS in molecular units by dividing TAS fluorescence intensity by the single transcript fluorescence units ( $q_{E1}$ ,  $q_{E2}$ , and  $q_I$ , respectively).

### **Splicing efficiency exhibits an ‘economy of scale’ behavior**

Having established the method, we next used it to determine how splicing efficiency changes with transcription level in individual cells. Quantifying splicing efficiency requires comparing the number of spliced transcripts to the total number of transcripts (spliced + unspliced) at the TAS. Here, the total transcript number was represented by  $N_{E1}$  or  $N_E$ , and the number of spliced transcripts was obtained by subtracting the number of pre-spliced transcripts  $N_I$ , from  $N_{E1}$  or  $N_{E2}$ . With these quantities, the splicing efficiency can be computed as  $\epsilon_i = (N_{Ei} - N_I)/N_{Ei} = 1 - N_I/N_{Ei}$ , where  $i=1$  or  $2$  denotes either of the two exon probe sets, and  $\epsilon_i$  should, in the absence of noise, be independent of  $i$ .

To cover a broad range of expression levels, we induced the Sonic Hedgehog (SHH) target gene *Gli1* in 3T3 cells with varying concentrations of recombinant SHH, and analyzed cells after 48h of induction<sup>31</sup>. In parallel, we analyzed the synthetic spliceable RG6 mini-gene described above, induced for 3h with a range of doxycycline concentrations. After induction, we fixed cells and performed smFISH-hybridization and imaging (SI Methods and Materials).

The Gli1 and CMV promoters expressed at up to ~20 transcripts per TAS. Altogether, we analyzed ~3000 (Gli1, Figure 4a) and ~1000 (RG6, Figure S5a) active sites in single cells, and computed the geometric mean for each condition (SI text), as well as the splicing efficiency in each cell (Figure 4b, and Figure S5b). For both genes, transcription level and splicing efficiency were heterogeneous, even at a single induction level<sup>18,32</sup>. This variability likely reflects both transcriptional bursting and other sources of biological variation<sup>32</sup>, as well as measurement errors from stochastic binding of probes and other sources (see SI text for more details).

While individual cells were variable, the mean splicing efficiency systematically increased with gene expression level. This ‘economy of scale’ behavior occurred for both genes. It was robust to experimental conditions, such as the strength and duration of induction. It was also robust to data analysis parameters (see SI text). Additionally, splicing efficiency increased with transcription level in a similar pattern for both genes (Figure 4c, 4d, 4e), reaching 80% of its maximum value at ~3.5 transcripts per TAS.

We ruled out potential artifacts that could appear to generate this ‘economy of scale’ behavior. For example, because imaging occurs at a fixed point in time, images can in general capture incomplete transcription events. If incomplete transcription were expression level-dependent, it would alter the two exon probe ratio,  $N_{E1}/N_{E2}$ , from its ideal value of 1. However, this ratio showed no systematic dependence on transcription level (Figure 4 and Figure S5). Additionally, to rule out potential misclassification of individual transcripts as the TAS, we used single-molecule DNA FISH to independently identify the TAS (Figure S6). Finally, analysis of two independent transcription level measurements ( $N_{E1}$  and  $N_{E2}$ ) enabled us to compute splicing efficiency and transcription rate using distinct exon readouts, avoiding a potentially spurious correlation between transcription level and splicing efficiency due to the appearance of  $N_{E1}$  in the expression for splicing efficiency,  $(1 - N_i/N_{E1})$ . (Note that without the second exon probe, hardness-ratios correction methods<sup>33,34</sup> from Astrophysics could also help address this issue (Figure S7 and SI text)). Together, these results support the validity of the economy of scale observation.

### **A mathematical model of ‘economy of scale’ splicing**

How can an enzymatic process such as splicing produce ‘economy of scale’ behavior? In the classic Michaelis-Menten model (Figure 5a and SI text) reaction efficiency declines monotonically with increasing substrate concentration (Figure 5b, black curve), producing the opposite ‘diminishing returns’ behavior.

One limitation of the simple Michaelis-Menten model is that it does not account for the inhomogeneous concentration of splicing machinery in dynamic interchromatin granule clusters called nuclear speckles<sup>35–37</sup>. Previous work has shown that more highly transcribed genes are closer to speckles<sup>36,38</sup>. To analyze the relationship between spatial organization and splicing efficiency, we analyzed splicing efficiency of RG6 as described above, and also performed

immunostaining to detect the splicing factor SC35 in the same cell (Figure 5c, SI Methods and Materials). We then quantified the distance from each TAS to its nearest speckle (Figure S8a and SI Text), and plotted the results as a function of splicing efficiency. This analysis revealed a correlation between splicing efficiency, transcription level, and the proximity of the TAS to the splicing speckle (Figure 5d and Figure S8b).

Together, these results suggest the hypothesis that stronger expression could increase the proximity of a gene to a speckle, which in turn could increase the availability of splicing machinery. To incorporate these effects into a modified version of the model, we allowed the rate of pre-mRNA binding to splicing machinery,  $k_{on}$ , to increase with the concentration of pre-mRNA (Figures 5. This simple modification generated economy of scale behavior at lower expression levels (Figure 5b, pink curve, and schemed in Figure S9a), switching to diminishing returns at higher expression levels as the splicing machinery eventually saturates (see additional mechanisms in the SI Text). These results show that a positive correlation between expression level and speckle accessibility could qualitatively explain ‘economy of scale’ splicing behavior.

## Discussion

Here, we introduced a quantitative imaging-based method to measure splicing efficiency in single cells, and used it to characterize the dependence of splicing efficiency on transcription level. It enables accurate intensity quantification of smFISH data to allow direct comparison of intensities of multiple channels at the same site. Although we focused on quantifying the splicing efficiency of constitutively spliced introns, the pipeline presented here can be extended to alternative splicing by incorporation of additional fluorescence channels.

The observed ‘economy of scale’ behavior is opposite to the ‘diminishing returns’ behavior one would expect from a standard enzymatic process. Mechanistically, it could reflect a disproportionate allocation of the shared splicing machinery ‘resource’ to more highly expressed genes. In fact, previous work has shown that different genes can effectively compete for splicing machinery inside the nucleus<sup>30</sup>, with more weakly expressed genes receiving less access to splicing factors<sup>36,38</sup>. This allocation of splicing resources could optimize the total amount of splicing that can be achieved by a fixed abundance of splicing components<sup>39</sup>.

Functionally, economy of scale acts as a non-linear filter within the overall gene regulation process, enhancing more strongly expressed genes (Figure 6a). In principle, this filter could impact many cellular regulation processes. For instance, it could help prevent pervasive low-level transcription<sup>40</sup> from inappropriately propagating to protein synthesis. It could also amplify, suppress, or reshape the mRNA distribution, depending on the underlying distribution



of transcription levels (Figure 6b and 6c). Thus, it could play an active role in controlling variation across a population of cells<sup>41,42</sup>.

There are several limitations of the present method. One is that it focuses on co-transcriptional splicing. Potential regulation through post-transcriptional splicing processes<sup>24,43</sup> is not captured. Second, because the method relies on direct binding of probes, occlusion of probe binding sites by bound proteins or secondary structure of the target RNA could in principle affect quantitation. Third, the post-splicing residence times for different products at the TAS are unknown. Their relative values affect the absolute magnitude of the measured efficiency, but not its dependence on transcription level. In principle, however, the dependence of splicing efficiency on transcription rate could be distorted if the residence times for different molecules depend in different ways on transcription rate (SI text).

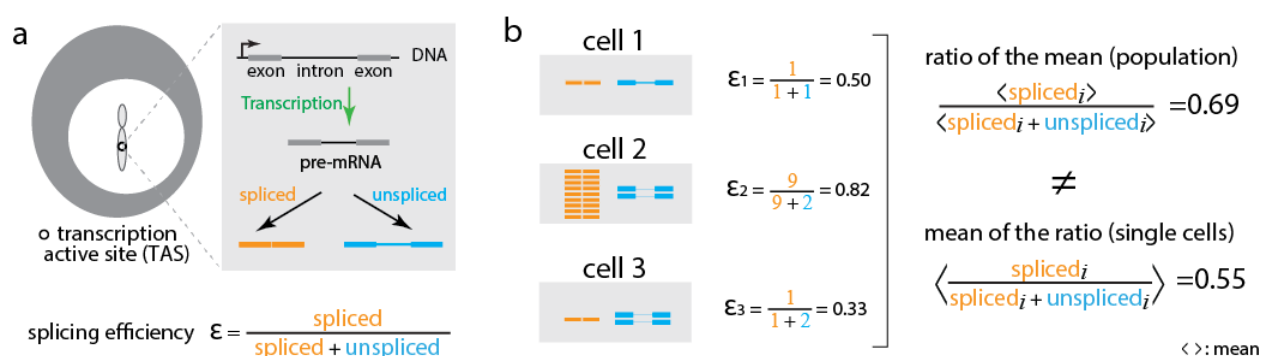
Because it operates quantitatively in single cells with subcellular resolution, this method should provide insight into kinetic features of the splicing mechanism. For example, by simultaneously imaging the spatial locations of splicing regulatory factors such as lncRNA alongside their target genes, it could enable one to determine how these factors affect splicing efficiency<sup>44–48</sup>. Using additional fluorescent channels, it could also allow analysis of correlations in splicing between neighboring genes, and enable comparison splicing efficiency between alleles of a gene within a single cell.

Further development of the method could address additional questions. For example, recent work has shown how comparison of nascent transcript levels with total transcript levels can provide information on dynamic changes in expression (RNA ‘velocities’) from single time-point snapshots<sup>49</sup>. These approaches could be combined with the analysis shown here to provide dynamic information on the relation of splicing to transcriptional bursting. Finally, by combining these approaches with sequential hybridization and barcoding techniques<sup>44,45,50</sup>, this method could enable genome-wide analysis of splicing efficiency in a single cell.

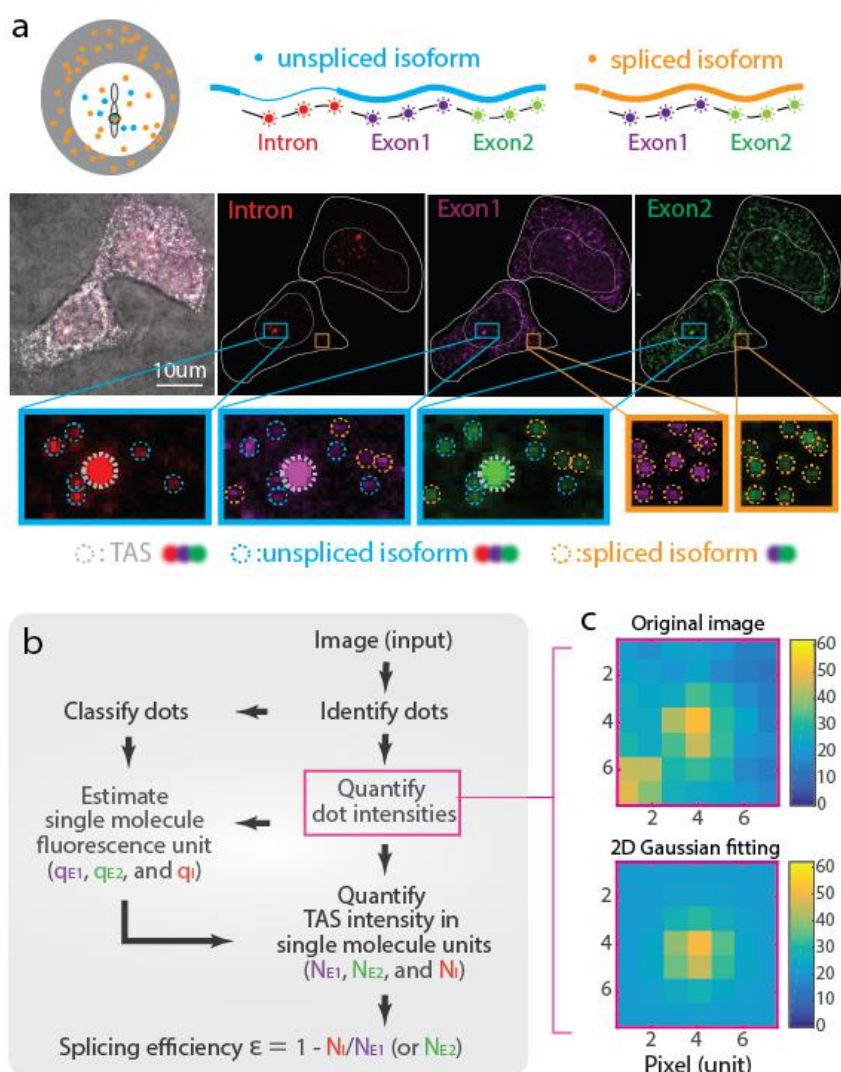
## Acknowledgements

The work was funded by a Fellowship from the Schlumberger Foundation, by the Gordon and Betty Moore Foundation through Grant GBMF2809 to the Caltech Programmable Molecular Technology Initiative and the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 from the U.S. Army Research Office. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred. M.B.E. is a Howard Hughes Medical Institute Investigator.

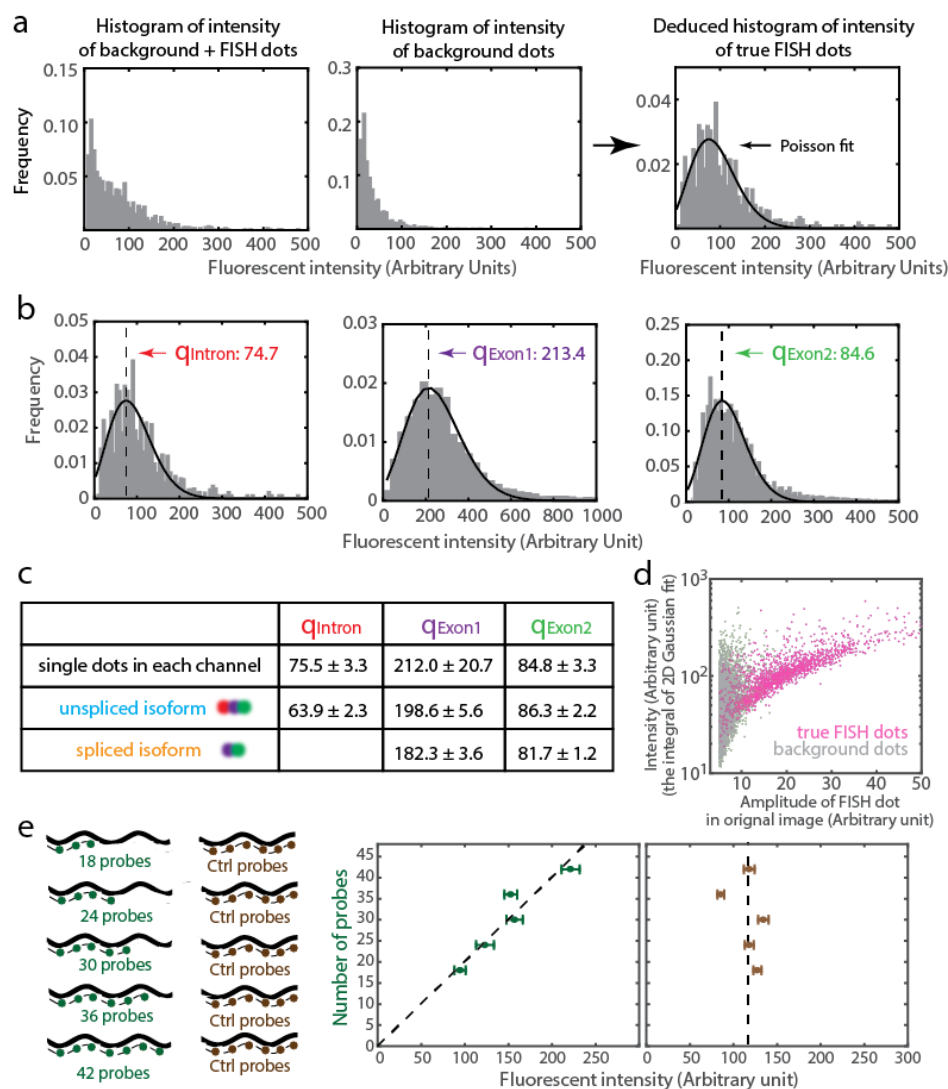




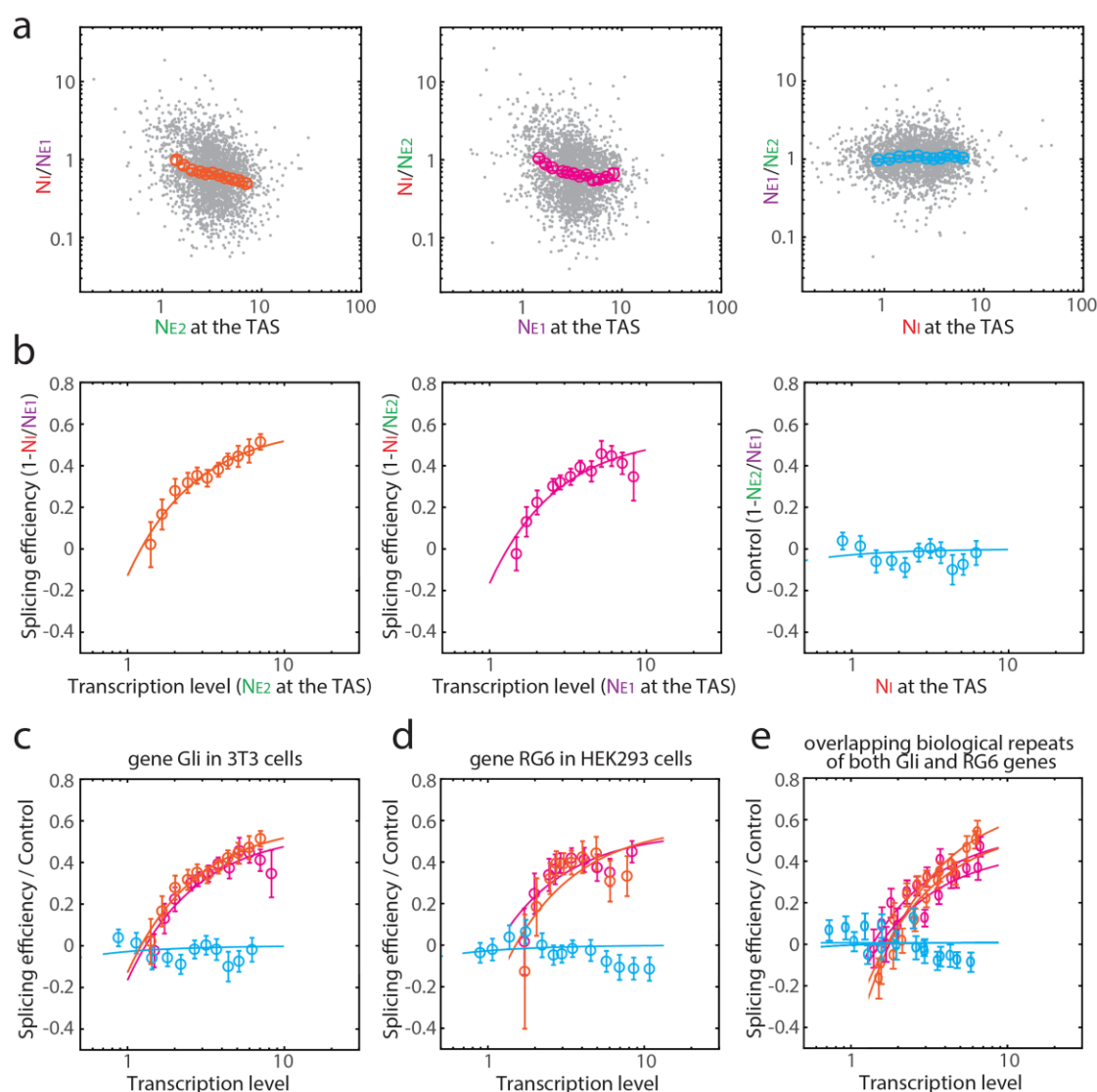
**Figure 1:** Splicing efficiency requires single cell measurement. (a) Splicing occurs co-transcriptionally at the TAS (white circle on chromosome). Each pre-mRNA molecule (gray) can be spliced to remove introns (orange) or left unspliced (blue). Splicing efficiency is the ratio between the number of spliced transcripts and the total number of transcripts at the TAS (equation, bottom). (b) Averaging over heterogeneous cells can distort splice efficiency. We consider three hypothetical cells with different numbers of spliced (orange) and unspliced (blue) transcripts. The mean of the splicing efficiency measured in each cell individually does not equal the ratio of the mean number of spliced transcripts and the mean number of total transcripts at the population level. Thus, splicing efficiency should ideally be measured in individual cells.



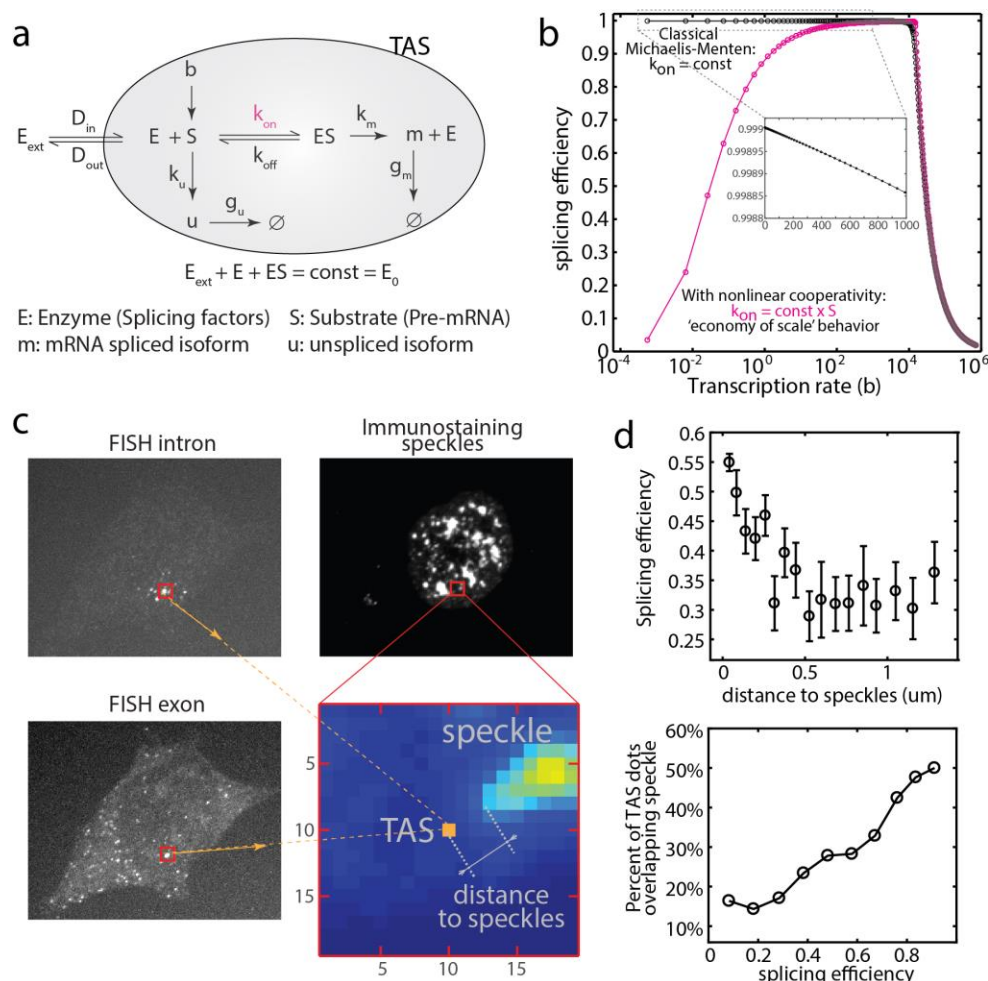
**Figure 2:** Experimental method for quantifying splicing efficiency at the TAS in individual cells. (a) We designed three smFISH probe sets targeting one spliceable intron (red), and two constitutive exons (purple, green). In this image of two cells, staining by each probe set is shown in the corresponding color channel, superimposed on a brightfield image (grayscale). Identified smFISH dots are circled in the zoomed insets, with TAS in gray, unspliced isoform (dots co-localized in Intron, Exon1, and Exon2 channels) in blue, and spliced isoform (dots co-localized in Exon1 and Exon2 channels) in orange. (b) Workflow for quantifying  $N_{E1}$ ,  $N_{E2}$  and  $N_I$ , the number of transcripts observed within the TAS in the exon1, exon2, and intron channels using corresponding single molecule fluorescence units,  $q_{E1}$ ,  $q_{E2}$ , and  $q_I$ , respectively. (See also Figure S1a.) (c) Iterative fitting of individual dot intensities. Using a method from stellar photometry of crowded star fields, we iteratively remove fluorescence from adjacent dots to produce a single dot image (bottom) from an initial image with multiple dots (top) (see SI Text and Figure S1c for more detail).



**Figure 3:** Quantification of single transcript intensity units in multiple channels. (a) The single dot transcript intensity distribution (right) can be obtained by subtracting a background distribution obtained without FISH probes (middle) from the foreground+background image obtained with FISH probes (left). (b) Applying this procedure to each channel, and fitting the resulting distributions to a Poisson distribution provides the three single transcript intensity units (indicated). (c) The measured single transcript intensity units in multiple channels are verified via dot-colocalization (Figure S1d). Different methods generate similar single molecule fluorescence units. Color is labeled as in Figure 2a. (d) The properties of background dots (gray) and identified true FISH dots (pink). Note overlap of these properties. See more dot fitting properties in Figure S2-S4. (e) The measured fluorescent intensity for staining of SDHA gene is proportional to the number of smFISH probes included (indicated numbers). A negative control (brown) uses a fixed number of probes, and displays a relatively fixed fluorescence intensity. This indicates that intensity scales linearly with number of probes.

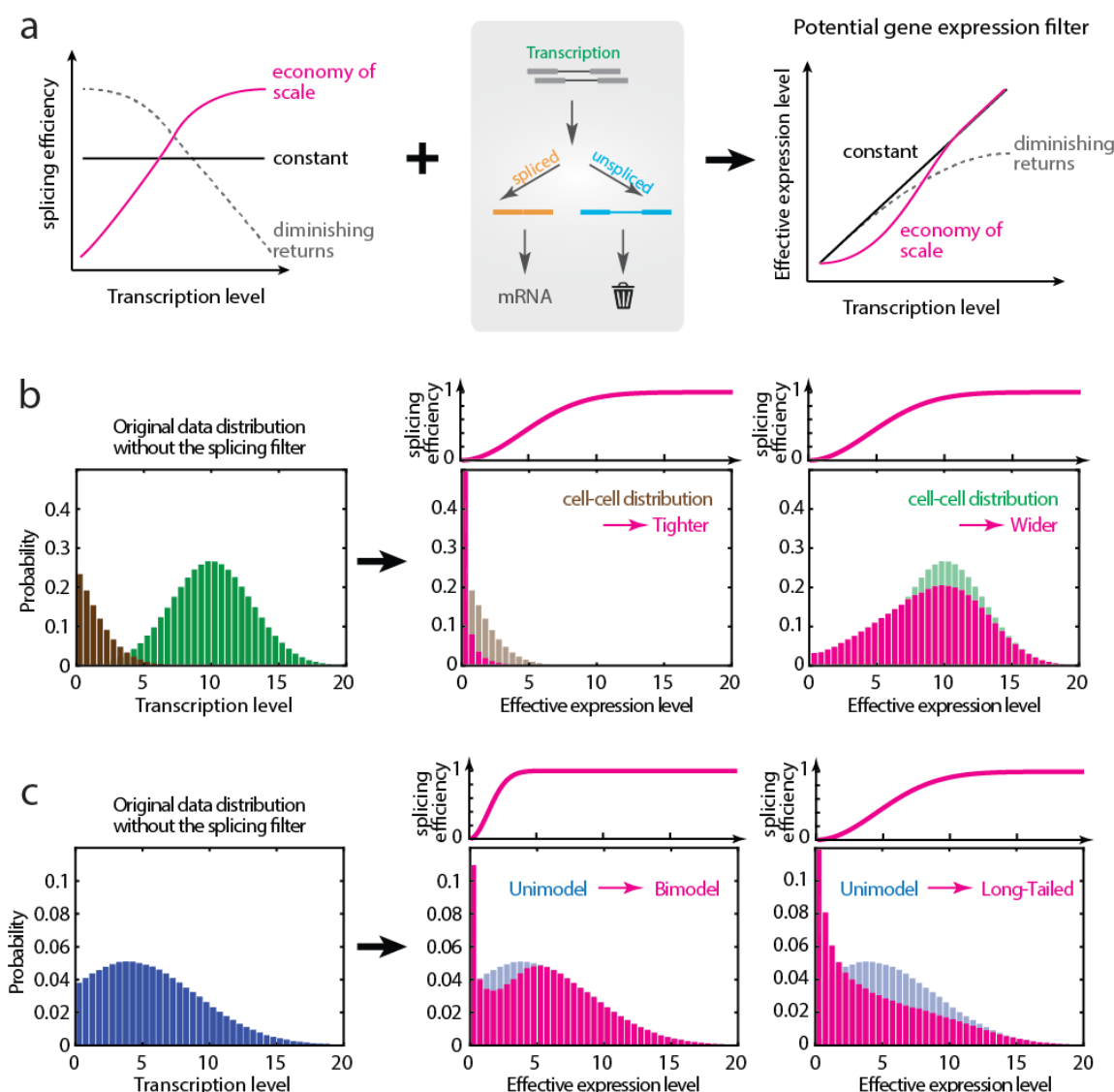


**Figure 4:** Splicing efficiency increases with transcript level, exhibiting ‘economy of scale’ behavior, for the two tested genes, Gli1 induced by Shh in 3T3 cells and the Tet-off mini-gene RG6 induced by dox in HEK293 cells. (a) Raw data of the number of transcripts measured by our methods. Each dot (gray) in the plot is one measurement from a single TAS. The number of TAS measured is ~3000 for Gli1 (and ~1000 for RG6 in Figure S5). Geometric means are shown in orange/pink/blue. Error bars represent standard error of the mean. (b) Based on (a), we find splicing efficiency ( $1 - N_i/N_{E1}$ , or  $1 - N_i/N_{E2}$ ) increases with transcription level (orange and pink), while control ( $1 - N_{E1}/N_{E2}$ , or  $1 - N_{E2}/N_{E1}$ ) measurements remained constant (blue). Solid lines are guides to the eye to highlight the ‘economy of scale’ behavior. The ‘economy of scale’ mathematical model is discussed in the main text and the corresponding Michaelis-Menten model fitting on the raw data were shown in Figure S9c. (c) Overlapped curves from (b). (d) Overlapped curves of the synthetic gene RG6. (e) Overlapped curves from biological repeats for Gli1 and RG6. All of them showed the repeatability of ‘economy of scale’ observation.



**Figure 5:** Potential model for 'economy of scale' behavior. (a) In a simple scheme based on the classical Michaelis-Menten model, splicing factors, collectively denoted E, bind to pre-mRNA, denoted S, at a rate  $k_{on}$  to form an enzyme-substrate complex, ES, which can unbind, at rate  $k_{off}$ , or transform into a mature mRNA, m, at rate  $k_m$ , releasing the enzyme.  $k_u$  denotes the production rate of unspliced RNA, denoted u,  $g_u$  and  $g_m$  denote the rate at which unspliced RNA or mRNA, respectively, are degraded or released from the TAS (shaded gray area).  $E_0$  denotes the total concentration of enzyme. (b) In the classical Michaelis-Menten model, splicing efficiency declines monotonically with transcription rate, in a "diminishing returns" manner (black). A variant of the model in which  $k_{on}$  is proportional to S, rather than constant, generates 'economy of scale' behavior (pink). Note that at higher transcription rates, splicing efficiency declines due to enzyme saturation. These 'diminishing returns' and 'economy of scale' behaviors were parameter independent (Figure S9b and SI text). (c) We measured the distance between TAS and its nearest speckle via smFISH and immunostaining in the same cell. FISH probes were designed as in Figure 2a. We used SC35 antibody to immunostain the speckles. The zoomed-in area was marked in red. smFISH defined the TAS location (orange), and immunostaining intensity showed the presence of speckles (Figure S8). (d) Splicing efficiency increases with proximity to speckles (top figure). TAS with higher splicing efficiency has greater probability of overlap with speckles (bottom figure). See raw data in Figure S8b.





**Figure 6:** Splicing can act as a gene expression ‘filter.’ (a) Different possible relationships between splicing efficiency and transcription level can generate different type of gene expression filter. ‘Economy of scale’ (pink) and ‘diminishing returns’ (gray) are high-pass and low-pass filters, respectively. (b) The ‘economy of scale’ filter can narrow or broaden the distribution of mRNA levels depending on the underlying distribution of mRNA expression levels. The filter (pink) preserves higher transcription levels (green) but suppresses lower transcription levels (brown). (c) Depending on its parameters, the ‘economy of scale’ filter can change the shape of an expression distribution from unimodal (blue) to bimodal (pink, center plot) or long-tailed (pink, right hand plot).

1. Bentley, D. L. Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* **15**, 163–175 (2014).
2. Das, R. *et al.* SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol. Cell* **26**, 867–881 (2007).
3. Rosonina, E. & Blencowe, B. J. Gene expression: the close coupling of transcription and splicing. *Curr. Biol.* **12**, R319–21 (2002).
4. Sharp, P. A. Splicing of messenger RNA precursors. *Science* **235**, 766–771 (1987).
5. Gilbert, W. Why genes in pieces? *Nature* **271**, 501 (1978).
6. Papasaikas, P. & Valcárcel, J. Evolution. Splicing in 4D. *Science* **338**, 1547–1548 (2012).
7. Wang, E. T. *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**, 470–476 (2008).
8. Houseley, J. & Tollervey, D. The many pathways of RNA degradation. *Cell* **136**, 763–776 (2009).
9. Boutz, P. L., Bhutkar, A. & Sharp, P. A. Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev.* **29**, 63–80 (2015).
10. Khodor, Y. L., Menet, J. S., Tolan, M. & Rosbash, M. Cotranscriptional splicing efficiency differs dramatically between *Drosophila* and mouse. *RNA* **18**, 2174–2186 (2012).
11. Khodor, Y. L. *et al.* Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in *Drosophila*. *Genes Dev.* **25**, 2502–2512 (2011).
12. Carrillo Oesterreich, F., Preibisch, S. & Neugebauer, K. M. Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Mol. Cell* **40**, 571–581 (2010).
13. Lipp, J. J., Marvin, M. C., Shokat, K. M. & Guthrie, C. SR protein kinases promote splicing of



- p>nonconsensus introns.
- Nat. Struct. Mol. Biol.*
- 22**
- , 611–617 (2015).
14. Katz, Y., Wang, E. T., Airolidi, E. M. & Burge, C. B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7**, 1009–1015 (2010).
15. Neugebauer, K. M. On the importance of being co-transcriptional. *J. Cell Sci.* **115**, 3865–3871 (2002).
16. Churchman, L. S. & Weissman, J. S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**, 368–373 (2011).
17. Herzel, L., Straube, K. & Neugebauer, K. M. Long-read sequencing of nascent RNA reveals coupling among RNA processing events. *Genome Res.* **28**, 1008–1019 (2018).
18. Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216–226 (2008).
19. Waks, Z., Klein, A. M. & Silver, P. A. Cell-to-cell variability of alternative RNA splicing. *Mol. Syst. Biol.* **7**, 506 (2011).
20. Larson, D. R. *et al.* Direct observation of frequency modulated transcription in single cells using light activation. *Elife* **2**, e00750 (2013).
21. Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* **4**, e309 (2006).
22. Vargas, D. Y. *et al.* Single-molecule imaging of transcriptionally coupled and uncoupled splicing. *Cell* **147**, 1054–1065 (2011).
23. Martin, R. M., Rino, J., Carvalho, C., Kirchhausen, T. & Carmo-Fonseca, M. Live-cell visualization of pre-mRNA splicing with single-molecule sensitivity. *Cell Rep.* **4**, 1144–1155 (2013).

24. Coulon, A. *et al.* Kinetic competition during the transcription cycle results in stochastic RNA processing. *Elife* **3**, (2014).
25. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
26. Zenklusen, D., Larson, D. R. & Singer, R. H. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* **15**, 1263–1271 (2008).
27. Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of single RNA transcripts in situ. *Science* **280**, 585–590 (1998).
28. Stetson, P. B. DAOPHOT - A computer program for crowded-field stellar photometry. *Publ. Astro. Soc. Pac.* **99**, 191 (1987).
29. Orengo, J. P., Bundman, D. & Cooper, T. A. A bichromatic fluorescent reporter for cell-based screens of alternative splicing. *Nucleic Acids Res.* **34**, e148 (2006).
30. Munding, E. M., Shiue, L., Katzman, S., Donohue, J. P. & Ares, M., Jr. Competition between pre-mRNAs for the splicing machinery drives global regulation of splicing. *Mol. Cell* **51**, 338–348 (2013).
31. Li, P. *et al.* Morphogen gradient reconstitution reveals Hedgehog pathway design principles. *Science* **360**, 543–548 (2018).
32. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
33. Park, T. *et al.* Bayesian Estimation of Hardness Ratios: Modeling and Computations. *Astrophys. J.* **652**, 610–628 (2006).

34. Coath, C. D., Steele, R. C. J. & Fred Lunnon, W. Statistical bias in isotope ratios. *J. Anal. At. Spectrom.* **28**, 52–58 (2013).
35. Lamond, A. I. & Spector, D. L. Nuclear speckles: a model for nuclear organelles. *Nat. Rev. Mol. Cell Biol.* **4**, 605–612 (2003).
36. Misteli, T., Cáceres, J. F. & Spector, D. L. The dynamics of a pre-mRNA splicing factor in living cells. *Nature* **387**, 523–527 (1997).
37. Phair, R. D. & Misteli, T. High mobility of proteins in the mammalian cell nucleus. *Nature* **404**, 604–609 (2000).
38. Quinodoz, S. A. *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174**, 744–757.e24 (2018).
39. Wahl, M. C., Will, C. L. & Lührmann, R. The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718 (2009).
40. Jensen, T. H., Jacquier, A. & Libri, D. Dealing with pervasive transcription. *Mol. Cell* **52**, 473–484 (2013).
41. Eldar, A. & Elowitz, M. B. Functional roles for noise in genetic circuits. *Nature* **467**, 167–173 (2010).
42. Losick, R. & Desplan, C. Stochasticity and Cell Fate. *Science* **320**, 65–68 (2008).
43. Girard, C. *et al.* Post-transcriptional spliceosomes are retained in nuclear speckles until splicing completion. *Nat. Commun.* **3**, 994 (2012).
44. Lubeck, E., Coskun, A. F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell in situ RNA profiling by sequential hybridization. *Nat. Methods* **11**, 360–361 (2014).
45. Shah, S., Lubeck, E., Zhou, W. & Cai, L. In Situ Transcription Profiling of Single Cells Reveals

- Spatial Organization of Cells in the Mouse Hippocampus. *Neuron* **92**, 342–357 (2016).
46. Zahler, A. M., Lane, W. S., Stolk, J. A. & Roth, M. B. SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev.* **6**, 837–847 (1992).
  47. Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R. & Misteli, T. Epigenetics in alternative pre-mRNA splicing. *Cell* **144**, 16–26 (2011).
  48. Engreitz, J. M. *et al.* Local regulation of gene expression by lncRNA promoters, transcription and splicing. *Nature* **539**, 452–455 (2016).
  49. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
  50. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, 6090 (2015).